DOCUMENT RESUME

ED 216 021 TM 820 169

AUTHOR McCann, Christina K.; And Others
TITLE The Evaluation of Middle School Programs.
PUB DATE [81]
NOTE 27p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Data Collection; *Educational Assessment; *Educational Objectives; Intermediate Grades; *Middle Schools; Program Effectiveness; *Program Evaluation

ABSTRACT

The applicability of conventional measurement techniques in the assessment and evaluation of a middle school educational program has been disputed. The objectives of this program include assisting students in the development of several areas. Among these are independence, healthy self concepts, viable ethical and value systems, socially responsible behavior, a knowledge base for continued learning, and appropriate interpersonal skills. Many of the institution's aspirations are in non-cognitive areas. Data for this evaluation should be collected on student performance in a natural setting. It should be multifaceted. This data may be generated from structured questioning which enables evaluators to draw conclusions about the effectiveness of program components. Multiple data sources relative to program expectations are a critical part of the question-directed evaluation model. Data sources should be valid, reliable, easily accessible, relatively inexpensive, and non-reactive. Two types of data sources are paper and pencil tests and unobtrusive measures. The various types of measures to be used are identified and explained. More evaluation studies on middle schools are needed to determine if major goals are being met, young adolescents are benefiting, and to understand methods of improving the middle school. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Mc Cann, C
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED216021

The Evaluation of Middle School Programs

J. Howard Johnston
Glenn C. Markle
University of Cincinnati

Christina K. McCann
University of Texas at Arlington

James Hogan
Forest Hills Local Schools
Cincinnati, Ohio

In the past, little research on middle schools, in particular their effectiveness has been conducted. Alexander (1969) attributed this lack of evaluation to people assuming that middle schools must be beneficial, a lack of appropriate measures, able evaluators and money.

Due to the increasing number of middle schools across the nation, attention towards their evaluation has grown. The major purposes for conducting middle school evaluations are to find out if the major goals of the middle school are being met, to find out if the young adolescents are benefiting and to understand methods of improving and strengthening the middle school (Alexander, et. al; 1969).

However, all program evaluation is designed to assist in making sound decisions about educational experiences (lewy, 1977). In a broad sense, according to Scriven (1967), evaluations' purpose is always the same: to determine the worth or value of something. If we can know that one program is more efficient than another, or that a program achieves or does not achieve its stated purposes, we are likely to make better decisions about the modification, maintenance or elimination of programs or program elements.

A variety of specific purposes have been suggested for program evaluations, but all are directed toward informing the curricular decision-making process. Doll (1974) lists two major reasons for conducting evaluations: to let participants in the program know what they have accomplished and to justify the expenditure of resources. Scriven (1967) views the process as more developmental and less political. He says that formative evalua-

TM 820 169

tion is done to improve the educational experience during its developmental stage, and summative evaluation is conducted to make an overall decision about the worth or value of an educational experience, such as in an end-of-term course evaluation. A distinguishing feature is that formative evaluation relies heavily upon the provision of feedback from the evaluation in a way that permits program modification.

David Payne (1974) synthesizes these views on the purposes of program evaluation when he proposes three general purposes:

(1) to improve the curriculum during its developmental stage,

(2) to allow for rational comparisons between competing programs, and

(3) to contribute to the general body of the knowledge about effective curriculum design.

Additional, more specific purposes for evaluation (Mehrens and Lehman, 1975; Ahman and Glock, 1975) are:

...to check school performance against community expectations,

...to make specific decisions about textbook adoption or the sequencing of content and learning experiences,

...to study articulation phenomena unique to the organizational pattern of a given district, and

...to identify the need for special programs or services within the system.

Curriculum evaluation can be formal or informal (Doll, 1974). In its most informal version, the evaluators are involved in "judging, estimating or giving opinions about the extent to which certain changes in school programs have occurred." Formal evaluations are characterized by the "Careful collection and treatment of data about progress toward prescribed goals" (Doll, 1974). In fact, all program evaluations involve some awareness of desired outcomes, whether formally stated or not, some type of "data"

collection, however informal, and a process of inferring and drawing conclusions about the program on the basis of the evidence at hand.

Initiating and Building Staff Support For Program Evaluation

When sufficient reason justifies a program evaluation, the possibility that the time, effort and cost are matched by the benefit received will depend on the reaction of the staff. The cardinal rule for building support for an evaluation program is to provide for the active involvement of all staff members in: 1) developing the plan for assessment design, 2) developing instruments for the evaluation, 3) analyzing evaluation results, and 4) implementing results in the curriculum (Fullan, Miles and Taylor, 1980).

Successful planning begins by making certain that the district, as represented by the superintendent, learns what it needs to know (Fullan, Miles and Taylor, 1980). Usually, the superintendent needs to know what is planned and should have the opportunity to identify any needs that he/she may have based upon contacts with the public at large and with school board members. The superintendent's perspective is unique and important and his/her support becomes especially important when the implementation phase of a plan is to begin.

Wisdom indicates that building principals should have an opportunity to contribute to the formulation of an evaluation plan; at the very least they should be able to critique the plan with constructive criticisms. Their special contribution is their knowledge about community or building characteristics which should be considered in the design of the evaluation. Also, beliefs held by principals should be known by the evaluator in advance so that the impact of the beliefs on the evaluation results can be hypothe-

sized and taken into account. Strengths and weaknesses of the variations introduced by the principals can be studied and accounted for. The evaluator is wise to involve the principal and consider his/her impact on the evaluation, the analysis of data and the development of implementation plans for he/she will most likely play the most significant role in the success or failure of the entire effort (Mehrens and Lehman, 1975).

It is the teachers, though, who are the foundation of success (Fullan, Miles, Taylor, 1980). Teachers play a critical and a key role in an evaluation effort. In most instances, teachers have knowledge or intuition about a myriad of factors affecting student behavior which need to be considered. They also are the persons most knowledgeable about the "true curriculum," that is the actual content of the curriculum as it is taught in the classroom. The teachers act as a guide to the design of an evaluation in harmony with realistic expectations (Odiorne, 1979). It should be added parenthetically that teachers, as well as the evaluation deisgned, will gain important insight into what is taught and how it is taught through participation in these early phases of an evaluation design.

The support for an evaluation effort and for use of the resultant data is easier to obtain when there is a long-standing practice of involving the staff in curriculum design and management (Ahman and Glock, 1975). If, however, the evaluation process or a high level of staff involvement in education decision-making are relatively new or undeveloped undertakings, there may be some skepticism about the intent of the evaluator. After all, teachers may ask, who or what is being evaluated? Confidence in the evaluator is fundamental for constructive staff participation in the evaluation process (Hersey and Blanchard, 1977).

The first task for staff involvement is an agreement on the scope and sequence of a curriculum. Here, curriculum is defined as the organized, teacher directed experiences provided for students. Deciding what the scope and sequence should be begins with the identification of a set of objectives that the staff can agree upon. Those lists can be proposed from a number of sources: district curriculum guides, commercial scope and sequence lists, professional literature, etc. Then by using any of several scales, the objectives can be ranked to determine those which receive priority attention for instruction. Sequencing the objectives evolves through group discussion about prerequisite student knowledge needed, student maturity, and materials available for teaching.

Once the scope and sequence is developed, it becomes possible to identify the items to be used to compose an evaluation instrument. Teachers should be extensively involved in the evaluation instrument development. It was pointed out earlier that teachers know what they expect students to be able to achieve and have models in their minds of the kind of information or concepts or skills that they expect their students to achieve. Teachers should be asked to submit questions in a format that can be quickly and easily scored and which they believe will test student achievement. The person responsible for the evaluation can pick and choose among the submitted questions and compose an evaluation instrument based upon his/her judgment about good items that represent the objective.

The completed instrument should be critiqued by the staff and the identified weaknesses resolved. Discussions among staff members should develop a concensus that the instrument will serve the purposes of the evaluation.

At this point, the evaluator needs to make some decisions about who can and should be involved in the review and evaluation process. Shall a committee representing the faculty be formed? Can the entire faculty be involved? Is the evaluator alone going to make the necessary decisions? Who will review and interpret the results of an evaluation? Who will develop a plan for using evaluation results in developing school programs? It is, of course recommended that broadly-based, reasonable and workable staff involvement be maintained.

Other decisions have to be made that should involve staff participation. As the evaluation plan takes shape, it is necessary to make the decision about which grade levels need to be involved. Should each grade level be involved? Should levels, such as primary, intermediate, middle and high school, be considered or should there be some other choices? The answer to this question should be guided by some beliefs about where effective changes can be brought about if change is necessary.

The results of program evaluation help to determine which objectives are being met and with what degree of success. It is thereby possible to make decisions about which objectives need to be emphasized, which objectives have been adequately achieved and which objectives can be relegated to a minor status. In other words, it is possible to identify objectives which are believed to be absolutely fundamental and those which are "nice to know."

Once the assessment instrument has been developed, the evaluator needs to determine when the assessment should be made. An advantage to giving the assessment at the beginning of a school year is that the benefits of the evaluation can be used during the remainder of the school year. Another

possibility is to assess a school program at the end of an academic year so that results can be associated with a milestone, and because the evaluator has the opportunity to use the summer months to process the data and prepare for fall activity. Regardless of the time table selected, it is important to get information back to those people who were involved in the developmental steps as soon as possible and in a form that is understandable to them.

Where objectives are met with a high level of success the most important effort remaining may be to try to determine if there is a sub-set of students who account for most of the errors made. If there is such a sub-set then adjustments in the instruction and curriculum for those students may account for all of the change the evaluator would care to make.

In those more likely circumstances where they are varying degrees of success with objectives, the administrators and the teachers need to be asked to help determine what the results mean. At this point, the evaluator may have his most difficult task because the effort must be kept constructive and positive. It is important to present the data in such a manner that it conveys an effort to describe status of pupil performance and to remove any hint of blaming or fault finding. Planning this phase commands the evaluator's attention for at this point the value of the effort can be gained or lost. The need is to analyze the results and to make whatever changes are necessary in the school programs.

A helpful technique may be to ask the teachers to hypothesize what they believe are the contributing factors to an unsatisfactory level of success. Brainstorming approaches applied to this problem can focus attention on a search for likely causes for test results. The evaluator's challenge is to keep the committee's efforts constructive and oriented toward problem-solving.

It should be remembered that it is important to keep the superintendent and district office personnel generally informed about the results of the evaluation. The support of these people will affirm a good problem-solving approach to the interpretation of the results by staff members and will be a positive factor in avoiding a blaming or negative attitude. The focus of effort, just as with principals and staff, should be upon the interpretation of results so that those persons involved have an opportunity to help find solutions.

Any of these new efforts can be used to help staff members to attain confidence in the evaluation and to support a plan to implement identified changes. It is through the development of a set of reasonable hypotheses about evaluation results that a unique opportunity becomes available to the staff, the evaluator and the district leadership to review curricular content, to reassess a commitment to the principles of child growth and development and to reinvestigate the effectiveness of current instructional strategies.

Given the diversity of teacher and administrator beliefs about the reasons for student performance there is likely to be some difficulty in achieving consensus about the identification of the causes of student performance. That is, brainstorming hypotheses about student performance may tend to become fault finding and "buck passing!" Perserverence in the effort to find the most plausable hypotheses is essential because the potential for effect on programs for students is at stake.

A consensus about the curricular implication is eventually woven into a plan for implementing an adjusted instructional program. Selection of texts, teaching aids, student schedules, student groupings, teacher assign-

ments, pre and post test materials, diagnostic tests, instructional approaches, and student schedules can be affected but eventually the plan for implementation must be generated if improvement and progress is to be achieved.

A Model For The Evaluation of School Programs

More and more professional educators have begun to question the applicability of conventional measurement techniques in the assessment and evaluation of middle school educational programs. The difficulty of employing conventional techniques, primarily paper and pencil tests, is magnified in the middle school because of the nature of that institution's objectives. A review of sample objectives advocated by Eichhorn (1966), Lounsbury and Vars (1978), Curtis and Bidwell (1977), Hansen and Hearn (1971), and Overly, Kinghorn and Preston (1972) illustrate this point most clearly. According to these writers, the middle school program should assist students in the development of...

...independence and self-reliance,

...healthy self concepts,

...viable ethical and values systems,

...a process orientation,

...adequate learning skills,

...self evaluation skills,

...socially responsible behavior,

...physical skills and appropriate health habits,

...a knowledge base for continued learning,

...curiosity, and

... appropriate interpersonal skills.

While some of the knowledge-related objectives stated above can be and probably should be assessed with conventional, valid and reliable measurement instruments, it is clear that a large portion of the institution's aspirations are in "non-cognitive" areas. Therefore, it is necessary to identify and develop valid and reliable methods for evaluating program effects in these less traditional learning areas. Failure to develop appropriate measurement techniques may lead to an over-reliance on traditional methods, regardless of their appropriateness, and to erroneous conclusions about the effects and effectiveness of middle school programs.

In addition to appropriateness, it is necessary to consider the "reactivity" of any given measure. Reactivity is the tendency for the measure to interfere with or influence the attitude or behavior being assessed. To illustrate, a basketball coach may include among the program's objectives the development of sportsmanship. On a paper and pencil test, the players may be able to identify the preferred responses and select those regardless of their personal attitudes on the topic. In this case, the paper and pencil test "interferes" with an accurate assessment of attitude and should be viewed with some suspicion. It may not indicate the program's effectiveness in producing desired attitudinal outcomes. A better or less reactive set of measures might include a tally of the number of personal fouls collected by the team, the number of technical fouls awarded the players (and coaches), and the number of on-court altercations which occur. These measures meet and serve to illustrate the criteria which are essential for program evaluation data.

1. The data should be collected on student performance in a natural setting.

2. The data should be multifaceted. That is, no single measure is used as a basis for judgment; several data sources are used.
3. The data should be non-reactive. The measurement of the attitude or behavior should not influence or interfere with the phenomena being studied.

Data which meet these criteria may be generated from a process of structured questioning which enables evaluators to draw conclusions about the effectiveness of a program's various components. By asking a carefully formulated and clearly defined series of questions, the evaluator will be led to the data sources which are most appropriate for the assessment of the specific program component under scrutiny. While the unique context of any given program evaluation will determine the precise nature of the questions asked, the following list serves as a general scope and sequence for a question-directed program evaluation effort:

1. What are we trying to evaluate? Precisely, what portion or component of the total educational "package" are we most interested in assessing? Program evaluation will be more effective if specific components of the program can be identified. To illustrate, it is more desirable to define the object of evaluation as the "intramural athletic program" than as "extra-curricular activities."
2. What do we expect this program component to accomplish? What are the specific objectives of the program and what do we expect students to do as a result of their participation in the program? For example, the intramural program might be expected to encourage students to develop an affinity toward participation in athletics as a recreational pursuit.

3. What will we accept as indicators that the program is achieving its objectives? Can we list the student behaviors that will allow us to conclude that the program element has had a desirable impact? Is our intramural program, for example, attracting a large number of students; do they participate in intramural-related activities outside of school; do they attempt to improve their performance through self-directed practice?
4. What sources of data are available which will indicate the presence or absence of the indicators? What kind of data on student performance do I have already, and what kind must be collected? Do we have access to student records; can we conduct structured observations; can we utilize student self-report data? In general, what kind of data can be accessed in a reasonable period of time with manageable effort?
5. What specific information from each source will be most valuable to our program evaluation? Within each data source what information is most clearly related to the objectives of the program? How can we gather, assemble and organize this information?
6. What does the data that we have collected tell us about our indicators? Does our data lead us to believe that the indicators are present or absent? What indicators are present?
7. Based on the presence or absence of our indicators, what can be concluded about the effectiveness of the program? Are the indicators of a program's effectiveness present in sufficient numbers and strength for us to conclude that our program is working?

Can we justify our conclusions on the basis of data rather than personal opinion or the strength of our desire for the program's success?

The organization of an inquiry such as that which is outlined above is as important as its content and sequence. In the absence of a clear organizational plan, potentially useful information might be overlooked or treated in a superficial manner. A model for the organization of the structured questioning approach to program evaluation is presented in Figure 1 as a guideline for managing the evaluation process. Just as the questioning approach can be modified to suit specific school settings, the organizational plan should be altered to accommodate the unique characteristics of the program being evaluated.

The foregoing evaluation process model is especially applicable to middle school settings where many of the goals and objectives are outside the domain of conventional assessment techniques. Failure to utilize assessment techniques which measure the unique objectives of the middle school will, inevitably, lead to erroneous conclusions about its effectiveness and potential for meeting the needs of young adolescents.

Instrumentation

Selecting the best available sources of data relative to program expectations is a critical part of the question-directed evaluation model. Multiple data sources are better than singular ones. That is, for each program expectation, student responses in a variety of settings rather than a single observation are more likely to result in valid decisions about the program. Data sources that are valid, reliable, easily accessible, relatively inexpensive, and non-reactive are most desirable. Two types of data sources are paper-pencil tests and unobtrusive measures.

What are we evaluating?	What do we expect?	What will indicate achievement?	What data sources are available?	Which will we use?	Results?
The eighth grade mathematics program.	Ability to solve word problems.	- solve word problems.	- paper & pencil test. - classroom discussion.	- paper & pencil test of ability to solve word problems.	
	Ability to perform mathematics	- solves equations correctly. - uses appropriate operations as dictated by notation. - avoids common errors.	- paper & pencil test. - observation during class. - homework review. - simulation.	- paper & pencil test of math operations.	
	Develops attitude toward mathematics and toward advanced study in mathematics.	- choose math over other subjects as "preferred". - use positive terms to describe mathematics. - agree with positive statements about math. - enroll in additional math courses. - attend classes. - read about math - choose math-oriented professions/car-er.	- subject preference survey. - enrollment data - adjective checklist - semantic differential. - Likert-scale - observation (contrived) - enroll. patterns. - attend. records. - library check-outs. - vocational preference inventory.	- subject preference survey. - semantic differential - enroll. patterns. - attendance.	

Paper-pencil tests can be standardized or locally produced. Regardless of the type, test items must reflect program objectives if the results are to serve as valid data sources for a program evaluation. Figure 1, for example, outlines an evaluation plan for an eighth grade mathematics program. A paper-pencil test is proposed as one data source to determine if the program develops the ability to solve word problems. Valid data will result only if the test actually measures the ability to solve word problems. It is not uncommon for an educational program to be criticized for the failure of its students to perform well on a test that measures outcomes unrelated to its objectives!

One technique to assure the validity of data sources is to match program objectives with test items. Subject area teachers can use course objectives as a basis for constructing test items. These items need to be reviewed by a panel of teachers to assure that they do indeed measure the objective, that they are clearly stated, and that there is general agreement on the correct response. The items generated by this process should be administered to a sample of students, perhaps one class, to determine estimates of item difficulty and discrimination. Good items, those that have a difficulty of approximately .3 to .8 and a discrimination index of .3 or higher, can then be selected for use on program evaluation tests. If a standardized test is used, a similar process should be followed, except for administering the test to a sample of students - the test company will have done this and should supply estimates of item difficulty and discrimination. The effectiveness of the program in achieving each expected outcome should be related to student performance on items specific to that expectation.

A semantic differential measures reactions to stimulus words and concepts in terms of ratings on bipolar scales. A series of scales attempts to assess a student's or a group's overall attitude regarding the concept. An example is:

"Mathematics"

good	____	X	____	____	____ bad
confusing	____	____	____	X	____ clear
strong	____	X	____	____	____ weak
boring	____	____	X	____	____ interesting

It is recommended that more than one but less than ten adjective pairs be used. Four pairs have been found to be adequate in most cases. When administering a semantic differential, direct students to indicate how they feel by checking the appropriate space between each pair of objectives and to record their first impression. To obtain a score, assign a numerical value (1-5) to each response and sum over all responses. In the above example, the score would be 15 (4 + 4 + 4 + 3).

A Likert scale measures students' opinions by soliciting the degree to which they agree with statements about the object, event or idea being assessed. Ten to twenty carefully selected statements usually provide sufficient data for estimating student attitudes. Individual responses are assigned a numerical value (usually 1-5) and a total score is computed by summing over all items. An example of a Likert scale to measure attitudes toward mathematics is:

1. I like mathematics.

   strongly agree agree uncertain disagree strongly disagree

2. It is important to be able to use skills taught in math class.

strongly agree agree uncertain disagree strongly disagree

3. I wish I didn't have to take courses in mathematics

strongly agree agree uncertain disagree strongly disagree

The score for the above responses would be 6 (2 + 3 + 1). Higher scores would indicate more positive attitudes.

Four general guidelines for constructing Likert scales are:

A. Understand clearly the object or its attribute about which opinions are to be solicited. Ambiguities appear in questions when the writer is unclear about the dimensions of the topic. The more specific attributes of a topic are susceptible to more precise opinion measurement. Thus, math class rather than the school and the school rather than the district should be the objects of question.

B. Be careful in the words used. Words selected should point directly to the object about which a response is desired.

C. Do not prepare more questions than can reasonably be used and the responses interpreted. Opinions are important and should be checked but the assessment should not be more complicated than the specific opinion warrants.

D. After preparing questions, let them sit for a few days; then revise them.

The Subject Preference Inventory is based upon the assumption that given a free choice between two subjects, students will select the one they like better. The inventory presents pairs of subjects to students who indicate a preference by circling their choice. Data from the Subject

Preference Inventory can be used to answer the question "Do students voluntarily select one subject over others more often after the course than before it?" A positive response is evidence of more favorable attitudes toward the course. A negative response indicates the need to examine course objectives and/or instructional strategies. An example of a limited subject preference survey is:

Math - Science	Literature - Science
History - Literature	History - Math
Science - History	Math - Literature

The score for each subject is the number of times the subject is circled. In above example, the science score is 3, the math score is 2, history is 1, and the literature score is 0.

The adjective checklist is a set of adjectives which students may circle to indicate their reaction. For example:

Math is...

___ informative	___ stimulating
___ boring	___ meaningless
___ beneficial	___ helpful
___ useful	___ dreaded

Responses are scored by summing the positive checks and subtracting negative ones. In the above example, the score is 2. Higher scores would indicate more positive attitudes.

"Reactivity" is a problem with many paper-pencil tests, especially those measuring affective outcomes. Students know, or can easily guess, the most desirable responses and answer the test items accordingly. Collecting responses in a non-threatening atmosphere and using anonymous

responses can help alleviate the problem. Unobtrusive measures are less reactive than paper-pencil tests and should be used as a second source of evaluative data.

Webb (1966) described five classes of information which lead to unobtrusive measures, i.e. measures that cause a minimum disruption to the variable being assessed. Digure 2 presents an outline of these five classes and includes various measures of desired outcomes within each class.

Physical traces refers to the accumulation or disappearance of "something" which is assumed to be related to the variable of interest. For example, if we wanted to know the popularity of a given television program, we could count the number of people watching the program. Such a "count" could, however, direct the interest of additional people to the program and therby alter the number of people watching. An unobtrusive measure was recently suggested by a worker at the Cincinnati Water Treatment Plant. He observed on a time-pressure chart that during the CBS Movie of the Week Airport, the water pressure dropped suddenly at the time of commercials. He inferred that the drop in pressure was due to flushing commodes and, after some computations, was able to estimate the number of flushes during commercials and thus, perhaps, the number of people watching the movie. By comparing drops in water pressures during subsequent programs he developed his own rating system! Such a measure is an example of physical erosion. Other, less bizarre physical traces include the number of math-related games students choose to use during a free period, the number of students who register for a free Math Occupations newsletter, and the number of students who go to the math lab during free periods. Such unobtrusive

measures should not be used alone but can add additional confidence to other measures of the same quantity.

Archives I and Archives II refer to written records that perserve data which can be used for numerous purposes. For example, school records of students enrolling in a specific course could be used as one source of data about student interest in that particular topic. The record of subjects which students elect may reflect changes in preferences over time. The type of library books or magazines most frequently read by students may be an indicator of attitudes.

Simple observations are observations of events which occur without special "external conditions", such as the observation of a student during a classroom discussion. A contrived observation is one where external conditions are manipulated. An example would be a classroom discussion during which a "smart" student, under the teacher's direction, proposes an incorrect solution to a math problem. The teacher then watches students to determine how many are willing to question the solution and to propose their own as an alternative.

Systematic record keeping of unobtrusive measures can reveal patterns of outcomes to guide the evaluation process. It cannot be emphasized too strongly that multiple measures must be employed; no one unobtrusive measure should be used to assess achievement of program objectives. However, when several bits of evidence point to the same conclusion, and evidence to the contrary is weak or absent, more confidence can be placed in the validity of our evaluations.

Figure 2: Unobtrusive Measures

Sources of Information

Method* / Indicator	Physical Traces Erosion and Accretion	Archives I: The running record	Archives II: episodic and private record	Simple Observation	Contrived Observation
Ability to solve word problems.		Scores on previous standardized tests that measure this skill.	Homework assign. successfully completed. Conversations with last year's math teacher.	Ability to solve word problems in a game designed for use during free time.	Ability to solve problems on blackboard during class time.
Ability to perform mathematical operations.		Check of previous math scores and math grades.	Homework performance.	Ability to keep score during games and to function effectively in situations requiring the use of addition, subtraction, etc.	Place student in charge of collecting money and reporting income from some school activity and check his performance.
Develops positive attitudes toward mathematics and toward advanced study in math.	Enrollment in subsequent math courses. Number of students who sign up to receive a free math newsletter.	Number of math-related books checked out of library. Number of math-related magazine articles read. Attendance in math class.	Number of math-related projects done in other courses. Number of inquiries in guidance office about math-related occupations or future courses.	Number of positive statements students make about math. Willingness to talk about math. Memberships in a math club.	Responses in a class discussion about the value of mathematics.

To better understand the worth and effectiveness of middle schools more evaluation studies need to be conducted. Suggested studies by Eichorn (1966), Alexander et. al. (1969) and Alexander and George (1981) focus on the young adolescent and the teacher of the young adolescent. Some of the suggested studies comparing and contrasting middle school setting versus a junior high school setting include: Student attitude, achievement, self concept, discipline, social acceptance, self directed learning, creativity, drop out rate, effects of a non-graded organization, teacher effectiveness, turnover rate, flexibility, participation and community involvement.

Selected References

Ahmann, S. and Glock, M. Evaluating Pupil Growth. (5th ed.) Boston: Allyn and Bacon, 1975.

Alexander, W., et. al. The Emergent Middle School. (2nd ed.) New York: Holt, Rinehart and Winston, 1969.

Alexander, W. and George. P., The Exemplary Middle School. New York: CBS College Publishing, 1981.

Curtis, Thomas E. and Wilma Bidwell. Curriculum and Instruction for Emerging Adolescence, Reading: Addison-Wesley, 1977.

Doll, Ronald C. Curriculum Improvement, Boston: Allyn and Bacon, 1974.

Eichorn, Donald H. The Middle School. New York: Center for Applied Research in Education, 1966.

Fullan, M., Miles, M. and Taylor, G. "Organization Development in Schools: The State of the Art," Review of Educational Research, 50:1:121-183, Spring, 1980.

Hansen, John H. and Arthur C. Hearn. The Middle School Program, Chicago: Rand-McNally, 1971.

Hersey, Paul and K.H. Blanchard. Management of Organizational Behavior: Utilizing Human Resources (3rd ed.), Englewood Cliffs: Prentice Hall, Inc., 1977.

Lewey, Arieh. Handbook of Curriculum Evaluation, New York: Longman, 1977.

Lounsbury, John H. and Gordon E. Vars. A Curriculum for the Middle School Years, New York: Harper and Row, 1978.

Odiorne, George. How Managers Make Things Happen, Englewood Cliffs: Prentice-Hall, 1979.

Overly, Donald E., John R. Kinghorn, and Richard L. Preston. The Middle School: Humanizing Education for Youth, Worthington, Ohio: Charles A. Jones Publishing Co., 1972.

Payne, David A. Curriculum Evaluation, Lexington: D.C. Heath, 1974.

Scriven, Michael. "The Methodology of Evaluation" in R.E. Tyler, ed., Perspectives of Curriculum Evaluation, Chicago: Rand-McNally, 1967.

Tittle, C., "Evaluative Studies", in Toward Adolescence: The Middle School Years. Seventy-Ninth Yearbook of the National Society for the Study of Education, Chicago: The University of Chicago Press, 1980.

Webb, Eugene J. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally, 1966.